09/830011 PCT/AU99/00914

Patent Office
Canberra

I, LEANNE MYNOTT, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PP 6606 for a patent by THE UNIVERSITY OF QUEENSLAND filed on 21 October 1998.

WITNESS my hand this
Twenty-ninth day of November 1999

LEANNE MYNOTT
TEAM LEADER EXAMINATION
SUPPORT AND SALES

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

P/00/009
Regulation 3.2

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION

Invention Title: "A METHOD OF PROTEIN ENGINEERING"

The invention is described in the following statement:

TITLE

"A METHOD OF PROTEIN ENGINEERING"

FIELD OF THE INVENTION

THIS INVENTION relates to a method for engineering proteins, and in particular, for engineering protein mimetics.

BACKGROUND OF THE INVENTION

Proteins are central to life, due to their crucial involvement in a variety of biological processes, such as enzyme catalysis of biochemical reactions, control of nucleic acid transcription and replication, hormonal regulation, signal transduction cascades and antigen recognition in the immune response.

"Protein" as used hereinafter, will include peptides, polypeptides and fragments thereof.

In many cases, one or more structural regions of a protein are responsible for a particular function, hereinafter referred to as "functional regions". These regions may constitute the active site of a protein enzyme, the nucleic acid binding domain of a transcription factor, region of a protein hormone crucial to binding the specific receptor for that hormone, or antigen-binding regions of antigen receptors.

A functional region of a protein usually comprises one or more amino acids which are required for that particular function, that is, they are essential for that function.

In many cases, although these required amino acid residues

are topographically proximal to each other, they may be well separated with respect to primary amino acid sequence, that is, they are non-contiguous. In addition, where there is more than one functional region of a protein, these regions may also be topographically proximal, but well separated in terms of primary amino acid sequence. In some cases, however, where there is more than one functional region involved in a particular function, these functional regions may also be topographically well separated. This is a particularly important point with regard to the functional regions of hormones.

"Hormone" as used hereinafter would typically be exemplified by soluble protein molecules such as interferons (e.g. IFN-γ), interleukins (e.g. IL-2), growth and differentiation factors (e.g. GM-CSF) and others such as growth hormone, prolactin, TGF-β, tumour necrosis factor and insulin. Each of these molecules is capable of binding a specific receptor and thereby eliciting a particular biological response or set of responses.

The fact that a particular function of a protein can be attributed to one or more functional regions of that protein has formed the basis for strategies aimed at modifying a protein by adding or subtracting functional regions to modify the function of that protein.

In this regard, the design and engineering of hormone mimetics has become an area of major importance, as many hormone-hormone receptor interactions are central to the regulation of a variety of

biological processes. It is envisaged that new mimetics will therefore become important new therapeutic agents that either mimic or inhibit the biological response to hormone-hormone receptor interactions.

A "mimetic" is a molecule which elicits a biological response either similar to, or more powerful than, that of the hormone (an "agonist"), or inhibits the action of the hormone (an "antagonist").

With regard to designing and engineering mimetics based on hormones, a problem frequently encountered with many engineered mimetics has been that they exhibit short biological half-lives and hence minimal bioavailability and efficacy. In this regard, it has been proposed that small cysteine-rich proteins might be useful as protein "scaffolds" as a basis for engineering mimetics, due to their stability (Vita *et al.*, 1995, Proc. Natl. Acad. Sci. USA **92** 6404). These small cysteine-rich proteins comprise a disulfide-bonded core and exposed amino acid side chains at the protein surface (Neilsen *et al.*, 1996, J. Mol. Biol. **263** 297). However the full potential of these proteins has not been realized due to the fact that typical prior art strategies for protein engineering have largely been limited to transferring or exchanging contiguous groups of amino acids within individual secondary structural elements, such as loops or helices or β-sheets.

Examples of such an approach would include: the exchange of secondary structural regions between RNase and angiogenin, either to confer RNase activity on angiogenin (Harper *et al.*, 1989, Biochemistry **28**

1875) or angiogenic activity on RNase (Raines *et al.*, 1995, J. Biol. Chem. **270**17180); the insertion of elastase inhibition activity into IL-1β by transfer of the protease inhibitor loop of elastase to the IL-1β scaffold (Wolfson *et al.*, 1993, Biochemistry **32** 5327); the insertion of a 10 amino acid calcium-binding loop of thermolysin into *Bacillus subtilis* neutral protease (Toma *et al.*, 1991, Biochemistry **30** 97); the insertion of a β–sheet from a snake toxin to replace the β–sheet of charybdotoxin (Drakopolou *et al.*, 1996, J. Biol. Chem. **271** 11979); and the incorporation of a β-sheet from carbonic anhydrase into the β-sheet of charybdotoxin (Pierret *et al.*, 1995, J. Med. Chem. **35** 2145).

Of growing importance in protein engineering has been the use of computer based technology combined with the elucidation of the 3D structures of small molecules and macromolecules. 3D molecular structures are being generated at an increasing rate, such as by X-Ray crystallography and NMR techniques. These 3D features can be stored in generally accessible, searchable databases, such as the BROOKHAVEN database.

For the purposes of this specification, a database will comprise a collection of "entries", each entry corresponding to a representation of an aspect of 3D structure of a framework protein. A framework protein is simply any protein for which a 3D structure exists, either by experimental elucidation or by predictive means such as computer modelling. A framework protein is potentially useful as a scaffold

which can be structurally modified for the purposes of imparting a particular function thereto.

A "query" will hereinafter refer to a representation of an aspect of 3D structure of a protein which exhibits a function of interest, hereinafter referred to as a "sample protein". The representation of 3D structure would be in a form suitable for searching a database with the intention of identifying a "hit". A hit is an entry identified according to the particular query and the algorithm used to perform the search.

An important advance in database searching has been made by representing 3D structures in terms of the relationship between atoms located in "distance space", rather than "Cartesian space" (Jakes & Willett, 1986, J. Mol. Graphics **4** 12; Ho & Marshall, 1993, J. Comp. Aided. Mol. Des. **7** 3). A location in Cartesian space is defined by three coordinates (x, y, z) which each correspond to a position along three respective axes (X, Y, Z), each axis being oriented at right angles to the other two.

A location in distance space, however, is defined by distances between atoms, expressed in the form of a distance matrix, which details the distance between atoms. Distance matrices are therefore coordinate independent, and comparisons between distance matrices can be made without restriction to a particular frame of reference, such as is required using Cartesian coordinates.

It is important to emphasise that an arrangement of atoms

and its mirror image are described by identical distance matrices. A root mean squared (RMS) difference can be used to alleviate this ambiguity.

With regard to the 3D structure of proteins, a simplification of protein structure can be provided by reducing a 3D structure to "Cα-Cβ vectors" as discussed in McKie *et al.*, 1995, Peptides: Chemistry, Structure & Biology p 354-355. A Cα-Cβ vector occupies a location in 3D space, the location being defined by the orientation of the covalent bond between the α carbon and β carbon atoms of an amino acid (Lauri & Bartlett, 1994, J. Comp. Aid. Mol. Des. **8** 51). It will be appreciated that each of the 20 naturally-occurring constituent amino acids of protein (except glycine), possess a Cα-Cβ vector due to the covalent bond between the "central" α carbon and the β carbon of the constituent side chain.

For those proteins containing Gly in the database, it is possible to mutate this to Ala to generate the required Cα-Cβ vector for database searching.

The usefulness of Cα-Cβ vectors is that they provide a simplification of 3D structure. Therefore, only the amino acid side-chains of a functional region of a protein need be represented by the Cα-Cβ vector map, thereby excluding the substantial portion of the protein(s) not directly involved in that particular function. For the purposes of database searching, Cα-Cβ vectors are ideal, as they constitute the basic 3D structural information needed.

After identification of Cα-Cβ vectors corresponding to a protein or a functional region thereof, the parameters that characterize each vector must be stored in a database in such a way that retrieval in response to a query can be made quickly. A number of options are available for suitable representation of Cα-Cβ vectors, whether as a database entry or as a query:-

(A) as a distance matrix;

(B) as a dihedral angle (δ) formed between respective Cα-Cβ vectors;

(C) as angles α_1 and α_2 formed between respective Cα-Cβ vectors.

A simple explanation of these representations is provided in Lauri & Bartlett, 1994, *supra*, which is hereinafter incorporated by reference. The key to successful database searching is speed and efficiency. Thus, computer search algorithms have been developed which use a strategy whereby the vast majority of entries in the database are eliminated in a preliminary screening step.

These algorithms are demanding of computer resources, and therefore a search is normally effected in two stages:-

(1) a screening search to eliminate entries that cannot possibly constitute a hit; and

(2) an atom-by-atom comparison of a query with each entry not eliminated in (1), to identify one or more

hits.

The search in (1) could screen entries based on geometric attributes of the query (Lesk, 1979, Commun. ACM **22** 219) interatomic distances and atom types (Jakes & Willett, 1986, *supra*), aromaticity, hybridization, connectivity, charge, position of lone pair electrons, or centre of mass of ring structures (Sheridan *et al.*, 1989, Proc. Natl Acad. Sci. **86** 8165). This screening process would eliminate entries that have no chance of meeting the 3D constraints of the query.

This strategy, although quick, requires that for an entry to register as a hit, it must comprise every specified query component. As the number of query components increases, the number of near misses increases and the likelihood of finding a hit decreases.

A more useful search strategy which assesses the relative merits of each near miss as well as each hit has recently been provided by the search program FOUNDATION (Ho & Marshall, 1993, *supra*). FOUNDATION uses a clique-detection algorithm (various algorithms are reviewed and compared in Brint & Willett, 1987, J. Mol. Graphics **5** 49 and Brint & Willett, 1987, Chem. Inf. Comput. Sci. **27** 152) which searches a 3D database of entries for a user-defined query consisting of the co-ordinates of various atoms and/or bonds of a 3D structural feature. FOUNDATION identifies all possible entries that contain any combination of a user-specified minimum number of matching atoms and/or bonds as hits.

Despite the usefulness of 3D database searching as a means of identifying structurally related proteins, this approach has not been well utilized with respect to engineering proteins with a desired function.

OBJECT OF THE INVENTION

The present inventors have recognized that 3D database searching is useful for identifying proteins which have one or more desired structural features, such proteins being suitable "frameworks" for the engineering of proteins with desired functions. Furthermore, the present inventors have realized that protein engineering is best achieved by modification of a framework protein to incorporate particular amino acid residues required for a function, rather than by incorporating entire elements of secondary structure such as loops or helices. This is particularly applicable when functionally important amino acids are scattered throughout a protein and are not confined to particular regions of secondary structure.

It is therefore an object of the present invention to provide a novel method of protein engineering.

SUMMARY OF THE INVENTION

The present invention resides in a method of protein engineering including the steps of:-

(i) creating a computer database which includes a plurality of entries, each said entry corresponding to a

description of a location and orientation in 3D space of the constituent amino acid residues of a framework protein, or a subgroup thereof;

(ii) creating a query corresponding to a description of a location and orientation in 3D space of two or more amino acid residues of a sample protein which are required for a function of said sample protein;

(iii) using said query and a computer program to search said database and thereby identify one or more hits; and

(iv) modifying an amino acid sequence of a framework protein which corresponds to a hit, by substituting one or more amino acid residues thereof with other amino acid residue(s) so that a framework protein with amino acid residues so modified is capable of exhibiting a function which is either similar to, or inhibitory of, a function of said sample protein.

Preferably, said location and orientation in 3D space is that of side-chains of the constituent amino acids of a framework protein.

Preferably, said location and orientation of each amino acid side-chain of said framework protein is simplified as a Cα-Cβ vector.

Alternatively, said aspect of 3D structure would be a conformation of a "backbone" of said framework protein. "Backbone" as

hereinafter defined, refers to carbon and nitrogen atoms and covalent bonds therebetween of amino acids of a protein, exclusive of amino acid side chain atoms.

Preferably, each said entry corresponds to a description in the form of a distance matrix representation of said Cα-Cβ vectors.

Alternatively, said Cα-Cβ vectors may be represented by dihedral angles or α_1 and α_2 angles.

Said framework protein could be any protein which exhibits one or more desired structural features which provide advantages in terms of size, solubility, resistance to enzyme digestion or the ability to retain structural integrity despite pH and redox changes.

Preferably, said framework protein is a small cysteine-rich protein capable of internal disulphide bond formation. Most preferably, said small cysteine-rich protein comprises less than 70 amino acid residues.

Suitably, said sample protein may be an enzyme, nucleic acid-binding protein, hormone, antigen, receptor, ion channel, chaperonin, or any protein with a function of interest.

Preferably, said sample protein is a hormone.

Preferably, said function of said sample protein comprises binding a specific receptor to thereby elicit a biological response. However, a variety of other functions are contemplated, such as catalysis, binding cations (Zn^{++}, Ca^{++}, Mg^{++}), transporting ions (e.g. Cl^-, K^+, Na^+),

binding lipids, binding nucleic acids as a means of transcriptional regulation or regulating DNA replication, assisting protein folding and transport, and any other function carried out by proteins.

Preferably, said location and orientation in 3D space is that of the side-chains of said two or more amino acid residues which are required for a function of said sample protein.

Preferably, said location and orientation in 3D space of each side-chain of said two or more amino acid residues required for said function is simplified as a Cα-Cβ vector.

Alternatively, a conformation of a backbone of said two or more amino acid residues which are required for said function may be applicable.

Preferably, said query corresponds to a description in the form of a distance matrix representation of Cα-Cβ vectors. However, other representations such as dihedral angles or α_1 and α_2 angles may also be applicable.

Preferably, said computer program used for searching said database incorporates the FOUNDATION algorithm (Ho & Marshall, 1993, *supra*, hereinafter incorporated by reference). Program FOUNDATION searches 3D databases of small organic molecules to identify structures that contain any combination of a user-specified minimum number of matching elements of a user-defined query. It achieves this by first using a distance matrix to define the topography of the query atoms, followed by

screening using various query constraints which define the chemical nature of the structure. The topology of the atoms in the structure are again represented using a distance matrix. Structural fragments in the database, whose distance description matches those of the query are identified using graph theory (Gibbons *Algorithmic Graph Theory*; Cambridge University Press: Cambridge, 1988).

In graph theory, a graph is a structure comprised of nodes (vertices) connected by edges. A graph is completely connected when all nodes are connected to one another. A subgraph is any subset of a larger graph. The largest completely connected subgraph of any graph is called a clique. Thus, the query is a completely connected graph, as all interatomic distances are determined in the distance matrix. The task is then to search a structural database to find all cliques that contain at least a user-defined number of matching nodes.

There are many clique-finding algorithms. Some of the well known procedures include those by Bonner, 1964, IBM J. Res. Develop., **8** 22; Gerhards & Lindenberg, 1981, Computing **27** 349 and Bron & Kerbosch, 1973, Commun. ACM **16** 575. Computational chemists have adapted these algorithms or implemented similar ideas to facilitate searching for 3D structures within databases (Kuntz *et al.*, 1982, J. Mol. Biol. **161** 269; DesJarlais *et al.*, 1988, J. Med. Chem. **31** 722; DesJarlais *et al.*, 1990, Proc. Natl. Acad. Sci. **87** 6644; Crandell & Smith, 1983, J. Chem. Infr. Comput. Sci. **23** 186; Brint & Willett, 1987, J. Mol. Graphics **5**

49-56; Kuhl *et al.*, 1984, J. Comput. Chem. **5** 24 and Smellie *et al.*, 1991, J. Chem. Inf. Sci. **31** 386).

The present inventors use a modified version of the clique-detection algorithm in program FOUNDATION as described by Ho & Marshall, 1993, J. Comp. Aided. Mol. Des. **7** 3-22. The search procedure is illustrated in Scheme A The major changes in comparison to Ho & Marshall, 1993, *supra* include:-

- the query and database structures are both proteins;
- the query elements are a distance matrix defining the topography of Cα-Cβ vectors, not individual atoms as in FOUNDATION;
- similarly, the structure is defined as a Cα-Cβ vector distance-matrix and not every atom as in FOUNDATION; and
- currently no query constraints are used to define the chemical nature of the structure.

An outline of the program we have written is shown in Scheme A. Prior to running the program, a particular sample protein is selected and the functional amino acids defined. The three dimensional structure of the sample protein must be determined by experimental or theoretical means well known in the art.

The first general step of the program is to read a set up file. This file includes the name of a file that contains a list of framework proteins to be searched, a name of a file that contains the co-ordinates of the functional epitope of the sample proteins, the name of the file that

contains the co-ordinates of the entire sample protein, required error ranges in Å for comparisons of Cα-Cβ vectors and the minimum number of Cα-Cβ vectors of the functional epitope required for a hit.

The next step is to calculate the distance matrix of the Cα-Cβ vectors of the query.

Each database file is now read in turn and the Cα-Cβ distance matrix of the framework protein is calculated. Using a series of automated scripts outlined in Scheme C, the database of small cysteine rich proteins is updated weekly by searching the Brookhaven database for suitable candidates. The clique detection algorithm of Ho & Marshall, 1993, *supra*, is used to identify geometric matches that match a user defined minimum number of query elements. Other algorithms well known to the art could be used at this stage. If no match is found, another database file is read and processed. If a Hit is found, its root mean square difference to the query is calculated, it is scored and results written to an output file. Another database file is read and the process repeated.

An alternative representation of the program is shown in Scheme B.

Alternatively, other applicable algorithms are provided by Brint & Willett, 1987, J. Mol. Graphics, *supra* and Brint & Willett, 1987, Chem. Inf. Comput. Sci. *supra*, which are hereinafter incorporated by reference.

In cases where backbone conformations are used as the

basis for database searching, applicable algorithms are provided by Holm & Sander, 1994, Proteins: Structure, Function and Genetics **19** 165; Airing, 1994, Corr. Open. Strict. Biol. **4** 429; Alexandrov *et al.*, 1996, Proteins: Structure, Function and Genetics **25** 354; and Alexandrov 1996, **9** 727, which are hereinafter incorporated by reference.

Suitably, said one or more hits correspond to respective entries identified by said algorithm according to said query.

Should there be more than one hit, it is desirable to evaluate and rank each hit. The most important factor in evaluating hits is "steric integrity", or the 3D structural complementarity of a hit when compared to a query. Several algorithms have been developed which could be utilized for this purpose. Such algorithms would include an algorithm used by the FOUNDATION program, algorithms which check van der Waals overlap of each said hit with said query (Allinger *et al.*, 1972, *supra*, hereinafter incorporated by reference), or algorithms which calculate volume in common and volume of extra space with respect to each said hit and said query (Marshall *et al.*, 1979, *supra*, hereinafter incorporated by reference).

It is also contemplated that other algorithms may be useful. For example, simple distance calculations between said hit and said query after superimposition thereof may be used to identify 3D spatial differences therebetween.

An outline of the process that is currently used for scoring is given in Scheme D. These scripts post process output data from the

database searching program, and components of these scripts will eventualy be incorporated into the program to provide a semi-automated process. In the current filtering process, steps 1, 2 and 3 address the potential for steric clashes with a receptor, steps 4 and 5 evaluate the conformational stability of the engineered hit, and step 6 provides optimization of the fit between a receptor and hit. Note that this filtering process is described with reference to scoring hits in terms of their predicted interaction with a receptor eg. a cytokine and cytokine receptor. One skilled in the art will realize that the principles outlined in Scheme D are applicable to any protein-protein interaction. For example, when a crystal structure is not known, scoring procedures can be implemented to ensure that the hit is subsumed by the steric surface of the ligand.

It is also envisaged that evaluation and ranking of each said hit may be achieved manually by a person skilled in the art, although this would be a less preferred method, particularly when there is a plurality of hits to be evaluated and ranked.

Preferably, an amino acid sequence of a framework protein which corresponds to a hit is modified by substituting one or more amino acid residues thereof with one or more amino acid residues selected from those which are required for a function of said sample protein. This engineering process can involve the addition, deletion and insertion of amino acids as desired.

It will be appreciated that the purpose of such modification is

to impart a particular function to a framework protein. The method of the invention takes account of the fact that the amino acid residues essential to a particular function will often be discontinuous with respect to primary sequence. These "scattered" amino acid residues may nevertheless form one or more functional regions, each of which occupies a distinct location and orientation in 3D space.

Modification of a framework protein will be performed so as to effectively "transfer" one or more functional region(s) to a framework protein. Transfer is achieved by incorporating amino acid residues from one or more functional regions of a sample protein into an amino acid sequence of a framework protein, said framework protein having been identified by a database search on the basis of desired 3D structural features. Such modification will be performed so as to engineer a protein which incorporates amino acid residues of said one or more functional region(s) appropriately located and oriented in 3D space.

Preferably, such modification will be performed so as to engineer a protein which incoporates amino acid residues of two or more functional region(s) of a sample protein.

Modification of a framework protein might be performed so that said framework protein is capable of exhibiting a function similar to that of said sample protein (such as in the case of an agonist), or alternatively, that inhibits a function of said sample protein (such as in the case of an antagonist).

Preferably, said framework protein would be modified so as to function as a mimetic.

However, the scope of the present invention extends to engineering proteins with any desired function by substituting amino acid residues of a framework protein. For example, an enzyme might be engineered to catalyze conversion of a substrate, or a transcription factor may be engineered to bind its cognate DNA sequence and to form complexes with other transcription factors necessary to promote transcription.

In the case where a mimetic is to be engineered, the preferred approach would be to modify an amino acid sequence of a framework protein (corresponding to a hit) by substituting amino acid residue(s) thereof with amino acid residue(s) of said hormone selected from those amino acid residues which are required for binding of said hormone to a specific receptor. Often, a biological response is elicited by a hormone binding to two or more receptor molecules, thereby cross-linking said receptor molecules. A hormone antagonist is therefore engineered by modifying a framework protein to include amino acid residues of a functional region required for binding one receptor molecule but not the other; an agonist is engineered by including amino acid residues of two functional regions, which together are required for binding and cross-linking of two receptor molecules. The functional regions required for binding said two receptor proteins would occupy unique

locations and orientations in 3D space. Engineering of an agonist would therefore require that the relative 3D location and orientation of each functional region is such that receptor binding and cross-linking is achievable.

An example of how a hormone agonist and antagonist can be so engineered will be discussed hereinafter.

Whichever approach is taken, modification of said amino acid sequence of said framework protein requires that considerations of maintaining stereochemical and secondary structural integrity apply. It is therefore important to be able to predict any structural effects induced in said framework protein by such modification. This can be accomplished with algorithms well known to the art as described in Bowie *et al.*, 1991, Science **253** 164-170; Luthy *et al.*, 1992, Nature **356** 83-85 and Laskowski *et al.*, 1993, J. Appl. Cryst. **26** 283-91.

Preferably, a modified framework protein would be chemically synthesized. Alternatively, this may be achieved by chemically synthesizing a polynucleotide sequence which encodes an amino acid sequence of said modified framework protein. Techniques applicable to the chemical synthesis of proteins and nucleic acids are well known in the art, and an example of such a technique will be provided hereinafter.

Alternatively, a polynucleotide sequence which encodes an amino acid sequence of a framework protein corresponding to said hit may be modified by *in vitro* mutagenesis techniques, resulting in a

modified polynucleotide sequence encoding an amino acid sequence of said modified framework protein. Suitable *in vitro* mutagenesis techniques are well known in the art, such as described in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Ausubel *et al.*, eds,) (John Wiley & Sons Inc 1995), which is hereinafter incorporated by reference.

According to one embodiment of the invention, each said entry in the database corresponds to a small cysteine-rich protein of no more than 70 amino acid residues, initially represented in cartesian coordinate form, but subsequently processed into a distance matrix representation of Cα-Cβ vectors prior to searching. Said query is in the form of a distance matrix representation of Cα-Cβ vectors corresponding to amino acid side-chains of said sample protein, said amino acid side-chains being required for high-affinity binding of said sample protein to a receptor protein.

Most preferably, said sample protein is human Growth Hormone (hGH), and said receptor protein is human Growth Hormone Receptor (hGHR). In this regard, said Cα-Cβ vectors of said sample protein are a simplification of the 3D location and orientation of the amino acid side-chains of hGH which contact hGHR during high-affinity binding, and are required for such binding.

~~In this case, said small cysteine-rich protein corresponding~~ to a hit is scyallotoxin, the amino acid sequence of which (shown in FIG. 1A) is modified so that a protein produced with that amino acid sequence

is capable of functioning as an hGH antagonist. The particular Cα-Cβ vectors used in the search process were Asp A171; Lys A172; Glu A174; Thr A175; Phe A176; Arg A178; Ile A179; Lys A41; Leu A45; Pro A48; Glu A56; Arg A64; and Gln A68. The particular amino acid residues of hGH incorporated into the amino acid sequence of scyallotoxin were selected from those required for high-affinity binding of hGH to hGHR (as shown above) and which conformationally matched with residues of scyallotoxin. Determination of which amino acids of scyallotoxin could be substituted without drastically affecting structural integrity was achieved with the assistance of the INSIGHT II modelling program.

Said hGH antagonist was chemically synthesised with the amino acid sequence shown in FIG. 1B.

In another case, said small cysteine-rich protein corresponding to a hit is a worm marine toxin (VIB). Said hit was identified by database searching using a query which comprised Cα-Cβ vectors of the following hGH amino acid residues: Lys A41; Leu A45; Pro A48; Glu A56; Arg A64; Gln A68; Asp A171; Lys A 172; Glu A174; Thr A175; Phe A 176; Arg A178; Ile A179; Arg A8; Leu A9; Asn A12; Leu A15; Arg A16; His A18; Arg A19; Tyr A103; Asp A116; Leu A117; Glu A119; and Thr A123.

An amino acid sequence of said hit is shown in FIG. 2A, and an amino acid sequence of a mimetic engineered by modifying one or more amino acids of said hit is shown in FIG. 2B. The particular amino acid residues of hGH used to modify said hit were selected from those

forming the agonist-binding functional region of hGH as indicated in FIG. 2. Overlap between hGH and said marine worm toxin is shown in FIG. 3, which serves to emphasize the ability of the method of the invention to identify hits which match cytokine agonist functional regions.

It is therefore expected that the engineered mimetic having the sequence shown in FIG. 2B will function as a hGH agonist.

A second aspect of the invention provides an engineered protein comprising a framework protein component and another protein component, wherein said framework protein component has 70 amino acid residues or less and has 2-11 disulfide bridges and said another protein component has at least one group of amino acids which are non-contiguous in primary sequence and which represent at least one functional region of said another protein component, whereby said engineered protein exhibits a function either similar to, or inhibitory of, said another protein component.

The framework protein component may comprise small disulfide rich peptides as described in U.S. Patent 5231011. However, it is desirable that the framework protein has particular characteristics such as rigidity, improved bioavailability, resistance to peptidases and proteases and stability to acidic conditions.

Preferably, said at least one group of amino acids represent at least two functional regions of said another protein component.

Most preferably, said at least one group of amino acids

represent two functional regions of said another protein component.

The said another protein component as well as having one group of amino acids which are non-contiguous in primary sequence may also have another group of amino acids which are contiguous in primary sequence.

Preferably, the engineered protein is selected from the group consisting of: a hGH anatagonist having an amino acid sequence as shown in FIG. 1B; and a hGH agonist having an amino acid sequence as shown in FIG. 2B.

It will be appreciated that according to both the first and second aspects of the invention, homologs of engineered proteins are contemplated. A person skilled in the art will realize that conservative amino acid substitutions, deletions and additions can be made such that a protein will retain a particular function notwithstanding such changes in amino acid sequence. All such homologs fall within the scope of the invention described herein.

EXPERIMENTAL

MATERIALS & METHODS

Synthesis of Growth Hormone Antagonist

A human growth hormone antagonist peptide was synthesized with the following sequence:-

(NH2-AFCNLRKCQDKCETFGLLGKCIGDKCECVK-OH).

The starting Resin for the synthesis was Boc Lys (ClZ)- Pam (A5J029 ABI 99.77%). Boc protected amino acids were purchased from NovaBiochem (Australia) or Bachem (Switzerland) The protecting groups used were Cystine (4-MeBzl), Lysine (Cl-Z), Aspartic acid (OcHex), Threonine (OBzl), Arginine (Tos), Glutamic acid (OBzl) and Asparagine (Xanthyl). Synthesis was carried out on a 0.1 mmol scale using a modified ABI 430A Peptide Synthesiser utilising HBTU *in situ* neutralisation for amino acid couplings. The protected amino acids (0.2 mmol) were predissolved in 4 mL HBTU/DMF (0.5 M) prior to loading on the synthesiser. Each amino acid was single coupled for 10 min.

After chain assembly, a sample of the peptide resin (439 mg) was treated with 9 mL Hydrogen Fluoride (HF) 500 mL p-cresol and 500 µL p-thiocresol for I hr at -5°C to 0°C. The HF was evacuated and the peptide triturated with 50 mL cold diethyl ether (Fluka) stirred for 2 min in an ice bath, then the precipitated peptide was isolated by filtration. The cleaved peptide resin was washed further with diethyl ether (50 mL) to remove residual scavenger, and then dissolved in 50% CH3CN/ H_20 with 0.1% TFA (100 mL). This solution was lyophilised to give 178 mg of crude peptide.

A sample of the crude peptide (60 mg) was purified by preparative RP-HPLC (Cl8 Vydac 218TP1022 250 X 22 mm id) using a linear gradient of 0-60% CH_3CN 0.1% TFA. The fractions were checked by mass spectrometry and those containing the peptide were combined

and the solution lyophilised overnight to give 14.72 mg of peptide. Mass of reduced MDPDM 1/10 MH+ 3354 calculated 3355.0.

A sample of the peptide (12 mg) was dissolved in 12 mL NH_4HCO_3 (0.I M) and stirred at room temperature overnight after which HPLC and mass spectrometry confirmed disulphide bond formation. The peptide solution was then purified by semi preparative RP-HPLC (CI8 Vydac 218TP1010 250 x 10 mm id) using a linear gradient of 0-60% CH_3CN 0.1% TFA. The purified product was relyophilised from H_20 three times and twice from 10 mM HCI before biological testing. Mass of oxidised MDPDM 1/10 MH+ 3348 calculated 3348.1.

^{1}H NMR spectroscopy

All NMR experiments were recorded on a Bruker ARX 500 spectrometer equipped with a Z-gradient unit. Peptide concentration was approximately 3 mM in 95% H_2O/5% D_20 (T = 293K). Spectra recorded included NOESY (Kumar *et al.*, 1980, Biochem. Biophys. Res. Comm. **95** 1; Jeener *et al.*, 1979, **71** 4546) with a mixing time of 400 millisecond, and TOCSY (Bax & Davis, 1985, **65** 355) with a mixing time of 85 millisecond. Spectra were run over 5550 Hz with 4K data points, 512 FIDs, 32-64 scans and a recycle delay of 1s. The solvent was suppressed using the WATERGATE sequence (Piotto *et al.*, J. Biomol. NMR, 1992, **2** 661) Spectra were processed using UXNMR. FIDS were multiplied by a polynomial function and apodised using a 90∞ shifted sine-bell function in both dimensions prior to Fourier transformation. Baseline correction using

a 5th order polynomial was applied and chemical shift values were referenced externally to DSS at 0.00 ppm. The random coil Hα chemical shift values of Wishart *et al.*, 1995, J. Biomol. NMR **6** 135, were used. Spectra were assigned using the methods of Wüthrich *et al.*, 1986, NMR of Proteins and Nucleic Acids. Wiley-Interscience NY.

Stability Tests

Blood Serum

Blood was collected in heparinised tubes by venapuncture. The blood was centrifuged at 5000 rpm for 20 min and the serum decanted. The blood serum was stored at -20°C. A sample of the blood serum (900 µL) was incubated with 100 µL of the stock peptide solution (I mg/mL in H_20) at 37°C and aliquots (100 µL) removed at the required time. A solution of 50% CH_3CN 0.1% TFA was added to precipitate the blood serum proteins and centrifuged at 13000 rpm for 5 min. A sample of this solution (100 µL) was analysed by RP-HPLC (Vydac C18 218TP54 250 x 4.1 mm id 1%/min gradient H_20/ CH_3CN 0.1% TFA) to detect peptide digestion.

Enzyme Stability Test.

Trypsin

To the peptide solution (NH_4HCO_3, pH 8.3, 0.87 mg/mL) was added trypsin (5% w:v). Samples were incubated at 37°C and aliquots removed at 0, 1, 3 and 18 hr and analysed by RP-HPLC as above.

α-Chymotrypsin

To the stock peptide solution (100 µL) was added 900 µL NH_4HCO_3 (pH 8.3). Chymotrypsin was added to 5% w:v and incubated at 37°C. Aliquots were removed at 0 hr, I hr and 24 hr and analysed by RP HPLC.

Pepsin

To the stock peptide solution (100 µL) was added H_20 (800 µL) and 0.1 M HCl (100 µL) to pH 2.2. Pepsin was added to give a 1% w:v solution and incubated at 37°C. Aliquots were removed at 0 hr, I hr and 24 hr and analysed by RP-HPLC.

Proliferation assay

BaF-B03 cells (a pro B cell line) that stably express the human Growth Hormone Receptor (hGHR) are used in this assay since they are able to elicit a GH-specific response at concentrations as low as 0.1 ng/mL hGH (4.54 pM). These cells also endogenously express the IL3 receptor and require IL3 or GM-CSF to survive in culture. The assay is based on that of Mossman, 1983, J. Immunol. Meth. **65** 55, and involves the following procedure:-

(i) culture cells in RPMI-1640 medium supplemented with 10% (v/v) foetal bovine serum (FBS) and 100 units/mL IL3 under 5% CO_2 at 37°C. Allow the culture to reach mid-log growth phase.

(ii) centrifuge cells at 500xg and wash with PBS to remove IL3 from the culture medium. Repeat the

centrifugation and resuspend in 1 mL of RPMI-1640 plus 0.5% (v/v) FBS. Count cells and dilute to a concentration of 8 x 10^5 cells/mL in same media.

(iii) from a constantly stirred suspension, add 50 µL of cells to each well of two 96 well plates.

(iv) prepare stock solutions of the mimetic to be tested at various concentrations such that the final concentration ranges from 100 nM to 100 µM made up in 0.5% FBS media (final volume is 150 µL, therefore stocks should be 3 times final concentration required). Add 50 µL of these solutions to cells in sextuplicate (i.e. A1 to A6 are identical etc.).

(v) prepare a stock solution (3 times) of hGH such that the final concentration is 0.5 ng/mL and add 50 mL to each well of one plate. Include one row as a negative control with no cytokine.

(vi) prepare a stock solution (3 times) of IL-3 such that the final concentration is 50 units/mL and add 50 µL to each well of the other plate. Include one row as a negative control with no cytokine.

(vii) incubate plates with no lids (to prevent uneven evaporation rates) in a vented humidified box under the abovementioned incubation conditions. Allow

incubation to continue for 24 hours.

(viii) add 50 µL of 4 mg/mL MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) to each well and incubate for a further 3 hours.

(ix) to stop assay, remove from incubator and lyse cells by adding 120 µL of isopropanol and triturating for several seconds per well or until cells are clearly lysed. Allow plate to rest in the dark for 5 minutes before reading.

(x) read plate at 595 nm on a microplate reader. Values obtained are directly proportional to cell number (as measured by mitochondrial dehydrogenase levels).

RESULTS AND DISCUSSION

1. *Overview of database search strategy*

A schematic of the computational approach we have developed is shown in FIG. 4. The first step involves the creation of a library of small cysteine-rich proteins. Currently, 344 such proteins (each with less than 70 amino acid residues) comprising over 3779 experimentally-derived 3D structures have been extracted from the BROOKHAVEN database. However, it would also be feasible to construct databases using theoretically derived features, such as by homology modelling, threading or other techniques known in the field.

Each structure has been simplified into Cα-Cβ vectors (step

a), and each used to build a database of entries (step b). For the purposes of searching the database, each query is in the form of a distance matrix representation of Cα-Cβ vectors (step c). However, it is possible to represent Cα-Cβ vectors by other means, such as dihedral angles (δ) or α_1 and α_2 angles. A simple description of these types of representations with respect to a Cα-Cβ vector pair is shown in FIG. 5.

The search algorithm compares the distance matrix representing the query Cα-Cβ vectors with the distance matrix representing Cα-Cβ vectors of each entry (step d). Comparison of conformational similarities was chosen because Cα-Cβ vectors are common to all amino acid side chains (except glycine), and are essentially anchored to the backbone. They therefore represent the initial orientation of the amino acid side chain in 3D space, which would probably not undergo significant change upon interaction with another protein. It is envisaged that the extra atoms of the side chain will provide some degree of induced fit during such an interaction.

Alternative, more restricted approaches would use secondary structural features such as α-carbon backbone structures, together with suitable algorithms well known in the field (Holm & Sander, 1994, *supra*; Alexandrov, 1996, *supra;* Alexandrov & Fisher, 1996, *supra*; and Oreng, 1994, *supra*).

The intermolecular geometric relationship of Cα-Cβ vectors is compared using the clique-detection algorithm of Ho & Marshall, 1993,

supra, which identifies hits according to a user-defined number of minimum vector components. However, other algorithms well known in the art would also be useful in this regard.

As a result of step d, one or more hits may be identified. If a single hit is obtained, no ranking is necessary. If the number of hits is small, it may be possible for the skilled person to evaluate and rank each hit individually (step e). If, however, the number of hits is large, such manual comparison would be more difficult, and an automated process is required.

The most important factor in evaluating and ranking hits is steric integrity, that is, the structural complementarity that each hit possesses with regard to the 3D space in which it must reside. For example, if the query is in the form of a distance matrix representation of Cα-Cβ vectors corresponding to the receptor-binding amino acid side-chains of a hormone, then a hit must be evaluated in terms of whether it would invade the 3D space accessed by the receptor upon binding the hormone. Several algorithms have been developed that are useful for this purpose. For example, the FOUNDATION program of Ho & Marshall, 1993, *supra* uses various flood filling algorithms to define the 3D space occupied by the receptor (as determined from the crystal structure of the receptor), and then uses atom-checking routines to establish whether the atoms of a hit reside in the binding "cavity" of the receptor. Other approaches include placing molecules in a cube containing lattice points

and checking the van der Waals overlap of each molecule (Allinger, 1972, In: Pharmacology and the future of Man. Proceedings of the 5th International Conference on Pharmacology pp 57-63). A related method involves the calculation of the volume in common and the volume of extra space of two molecules (Marshall *et al.*, 1979, The Conformational Parameter in Drug Design: The active analog approach. **112** 205).

It is also possible to use simple distance calculations between query and hit, after the two have been superimposed, to identify if the hit protrudes from the space occupied by the query structure. This is an approach the present inventors have implemented in an algorithm currently being constructed.

It is also important to be able to predict any drastic structural effects that may result from amino acid sequence changes when modifying a hit. This will, in part, be achieved by maximizing the degree of amino acid sequence identity of the modified hit with that of the protein (or area of the protein) to which the query corresponded. In addition, the stereochemical and degree of secondary structure disruption of the modified hit can be evaluated using standard algorithms which check protein stereochemistry on an amino acid by amino acid basis. Similarly, secondary structure prediction algorithms can be used to evaluate the potential for an amino acid sequence modification of a hit to disrupt secondary structure.

Finally, the present inventors plan to utilize molecular

surfaces to compare various physicochemical properties of a query and hit. Charge, electrostatic potential, hydrophobicity, occupancy, and hydrogen bonding potential have all been mapped to protein surfaces, providing detailed comparisons between proteins. A method for quantitating the degree of similarity between two molecular surfaces has been developed, in which a gnomonic projection casts the calculated values of a given property onto a spherical surface (Dasnzinger & Dean, 1985, J. Theor. Biol. **116** 215). Two such surfaces can then be superimposed using pairs of corresponding atoms. This algorithm would be very useful for comparing query protein with a hit, to allow fine tuning of amino acid residues of the protein corresponding to the hit, and to improve steric and electrochemical complementarity.

Since the database searching algorithms (such as provided by the FOUNDATION program) applicable to the method of the invention allow for the identification of partial hits, there is scope for a skilled person to use molecular modelling to identify additional regions on the surface of the protein corresponding to the partial hit for mimicking vectors missed in the database search. This could involve the use of D-amino acids or non-coded amino acids, for example, to achieve better mimicry when engineering a mimetic.

2. *Engineering a Growth Hormone Mimetic*

Growth hormone (GH) is a pituitary hormone that regulates many growth processes, such as the growth and differentiation of muscle,

bone and cartilage cells. The growth hormone receptor (GHR) consists of three domains: -

(i) an extracellular domain that binds GH;

(ii) a transmembrane domain; and

(iii) a cytoplasmic domain involved in eliciting an intracellular signal upon hormone binding.

Intracellular signalling occurs as a result of dimerization of separate GHRs following sequential binding of each receptor to a single GH ligand. The first GHR binds to the high affinity site of GH, while the second GHR subsequently binds binding to this complex. In support of this model, the crystal structure of this complex shows two identical receptor molecules bound to dissimilar sites on a single human GH molecule (hGH; De Vos *et al.*, 1992, Science **255** 306).

The high affinity site on hGH is concave and buries approximately 1200 A^2 of surface area, while the second binding site on hGH buries approximately 900 A^2 of surface area. A third region contributing to the stability of the complex comprises an area of 500 A^2 buried by the receptor-receptor interaction.

The crystal structure also reveals that the actual contact areas of both the high affinity and low affinity sites of hGH are buried upon complexation with the receptors.

In developing antagonists of hGH, the present inventors have sought to design molecules that mimic the high-affinity binding of

hGH. Mutagenic studies of the amino acid residues within the high affinity binding site showed a dramatic decrease in affinity when certain of these amino acid residues were converted to alanine (Cunningham & Wells, 1993, **234** 554). In this regard, of the 31 amino acid residues with buried side-chains, a mere eight (Lys A41; Lys A45; Pro A61; Arg A64; Lys A172; Thr A175; Phe A176; and Arg A178) accounted for approximately 85% of the total change in binding energy resulting from substitution by alanine. A further five residues (Pro A48; Glu A56; Gln A68; Asp A171; and Ile A179) essentially accounted for the remainder of the binding energy.

The GH residues currently used in the design of antagonists are: Asp A171; Lys A172; Glu A174; Thr A175; Phe A176; Arg A178; Ile A179; Lys A41; Leu A45; Pro A48; Glu A56; Arg A64; and Gln A68. It is these amino acid residues of hGH which formed the basis of the query for the purposes of database searching.

The highest-ranked hit obtained as a result of the search is illustrated in FIG. 6. The hit corresponded to a thirty amino acid residue scorpion toxin (scyllatoxin) protein which presents almost identical Cα-Cβ vectors to that of the buried contact area of hGH. The RMS difference of the respective Cα-Cβ vectors is 0.07.

The amino acid sequence of scyllatoxin is shown in FIG. 1A, and the sequence of the hGH mimetic engineered from this sequence is shown in FIG. 1B. Molecular modelling studies (using INSIGHT II)

suggested that the C-terminal His of the scyallotoxin-based mimetic could be removed with no effect on biological activity. This residue was removed to prevent racemization of the His. The amino acid sequence in FIG. 1B was prepared synthetically, purified and characterized as described in ***MATERIALS AND METHODS***.

3. *NMR spectroscopic analysis*

Due to difficulties in digesting the material to elucidate disulphide bond formation, NMR was used for this analysis. Secondary Hα shifts for scyllatoxin and the hGH mimetic are shown in FIG. 7, which provides strong evidence that the hGH mimetic and scyllatoxin share similar secondary structure, and importantly, a similar arrangement of disulphide bonds. The presence of such disulphide bonds is crucial for the improved stability of a mimetic based on a small cysteine-rich protein.

A schematic representation of the secondary structure of the hGH mimetic deduced by NMR is shown in FIG. 8. Initial analysis of NOEs suggested that the hGH antagonist forms the following structures: residues 1-5, extended; residues 6-16, α-helix (regular); residues 16-18, β-turn; residues 18-22; β-sheet; residues 22-25, β-turn; residues 25-30, β-sheet. These last three elements of secondary structure are combined to form a β-hairpin. Identical elements of secondary structure are found in scyallotoxin.

4. *Bioactivity*

The hGH mimetic was tested for biological function by

bioassay using the BaF3 cell line. The results are shown in FIG. 9. The hGH mimetic was assayed at various concentrations to check its ability to inhibit BaF3 cell proliferation in response to either 0.5 ng/mL hGH, or as a control, 50 Units/mL IL-3. The calculated K_i from these experiments was approximately 200 μM, and no inhibitory activity was observed with respect to IL-3 induced proliferation. Thus, the engineered hGH mimetic was a specific antagonist of hGH.

5. *Bioavailability*

Preliminary studies evaluated the bioavailability of the hGH mimetic by exposing it to a variety of proteases (trypsin, α chymotrypsin and pepsin) and blood serum proteins as described in ***MATERIALS AND METHODS***. The results of the blood serum stability test are presented in Table 1, and the results of the enzyme stability tests are presented in Table 2. The hGH mimetic was found to be stable after 24 hrs in each case, while control peptides were rapidly digested.

CONCLUSION

These studies have shown that by engineering small, cysteine-rich proteins, a stable mimetic with high bioavailability can be made with desired biological characteristics, in this case the ability to antagonize the biological action of hGH. Furthermore, the database searching strategy of the present invention has shown that suitable "scaffolds" for engineering mimetics can be identified according to aspects of 3D structure which are shared with a sample protein that possesses a

function of interest. Finally, once suitable scaffolds are identified, modification of the amino acid sequence of the scaffold can be performed so as to impart a function of the sample protein, or a function antagonistic thereto.

The present invention therefore provides a new strategy for the engineering of proteins, which strategy is particularly applicable to the engineering of mimetics which may constitute the next generation of therapeutics.

TABLES

TABLE 1 Blood serum stability test results

	0 hr	1 hr	24 hr
Control peptide	partially digested after 3 min	fully digested	
hGH mimetic	stable	stable	stable

TABLE 2 Enzyme stability test results

	Control peptide	hGH mimetic
trypsin	Digested in 1 hr	Stable over 18 hr
α-chymotrypsin	Digested in 1 hr	Stable over 18 hr
pepsin	Digested in 1 hr	Stable over 18 hr

LEGENDS

TABLE 1

hGH mimetic was a solution of MDPDM 1/10

TABLE 2

hGH mimetic was a solution of MDPDM 1/10

FIG. 1

Amino acid sequences of scyallotoxin (A), and the engineered hGH antagonist (B).

FIG. 2

Amino acid sequences of (A) marine worm venom (VIB) and (B) hGH agonist. Also indicated are the particular hGH amino acid residues used to modify the sequence in (A) and thereby arrive at the engineered hGH agonist shown in (B).

FIG. 3

Overlay of backbone structures of hGH and marine worm toxin.

FIG. 4

Schematic overview of database searching strategy.

FIG. 5

Two-dimensional depiction of three different representations of a pair of Cα-Cβ vectors: d = interatomic distance as used to construct distance matrices; δ = dihedral angle; α_1 and α_2 angles.

FIG. 6

Overlay of Cα-Cβ vectors corresponding to hGH (grey) and scyallotoxin

(black), showing substantial identity therebetween.

FIG. 7

Comparison of secondary Hα shifts for scyallotoxin and the hGH antagonist showing substantially identical structure and disulphide connectivities.

FIG. 8

Schematic depiction of secondary structure of hGH antagonist based on results of NMR analysis.

FIG. 9

Specific effect of hGH antagonist on BaF3 cell proliferation by inhibiting the growth response of the cells to 0.5 ng/mL hGH, but not to 50 U/mL IL-3.

DATED this Twenty-first day of October 1998.

THE UNIVERSITY OF QUEENSLAND,

by their Patent Attorneys,

FISHER ADAMS KELLY.

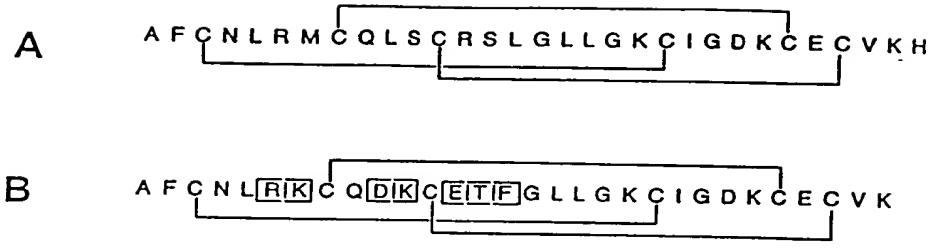
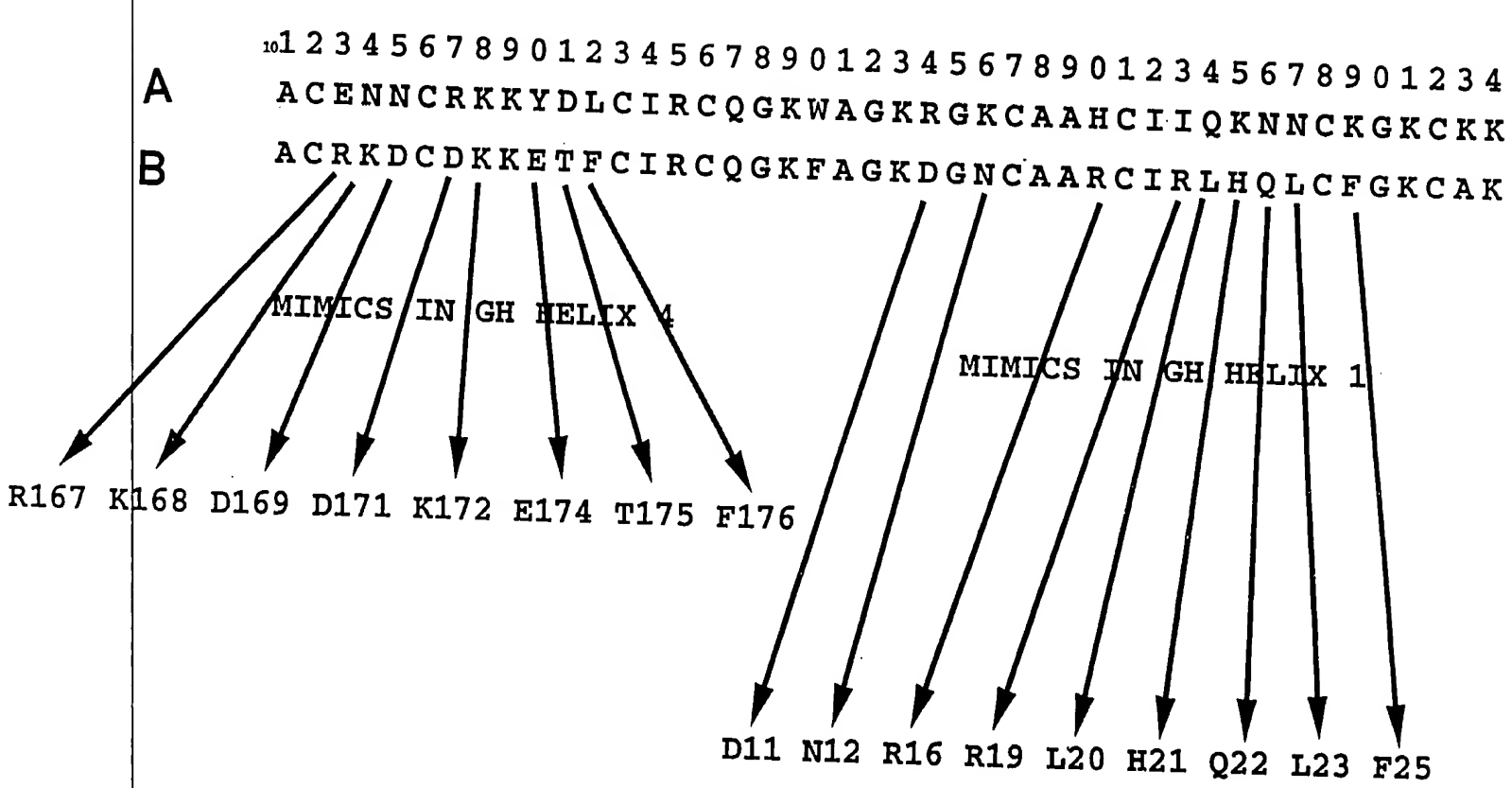

A
A F C N L R M C Q L S C R S L G L L G K C I G D K C E C V K H
B
A F C N L R K C Q D K C E T F G L L G K C I G D K C E C V K

Figure 1

FIGURE 2


10 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4
A ACENNCRKKYDLCIRCQGKWAGKRGKCAAHCIIQKNNCKGKCKK
B ACRKDCDKKETFCIRCQGKFAGKDGNCAARCIRLHQLCFGKCAK
MIMICS IN GH HELIX 4
MIMICS IN GH HELIX 1
R167 K168 D169 D171 K172 E174 T175 F176
D11 N12 R16 R19 L20 H21 Q22 L23 F25

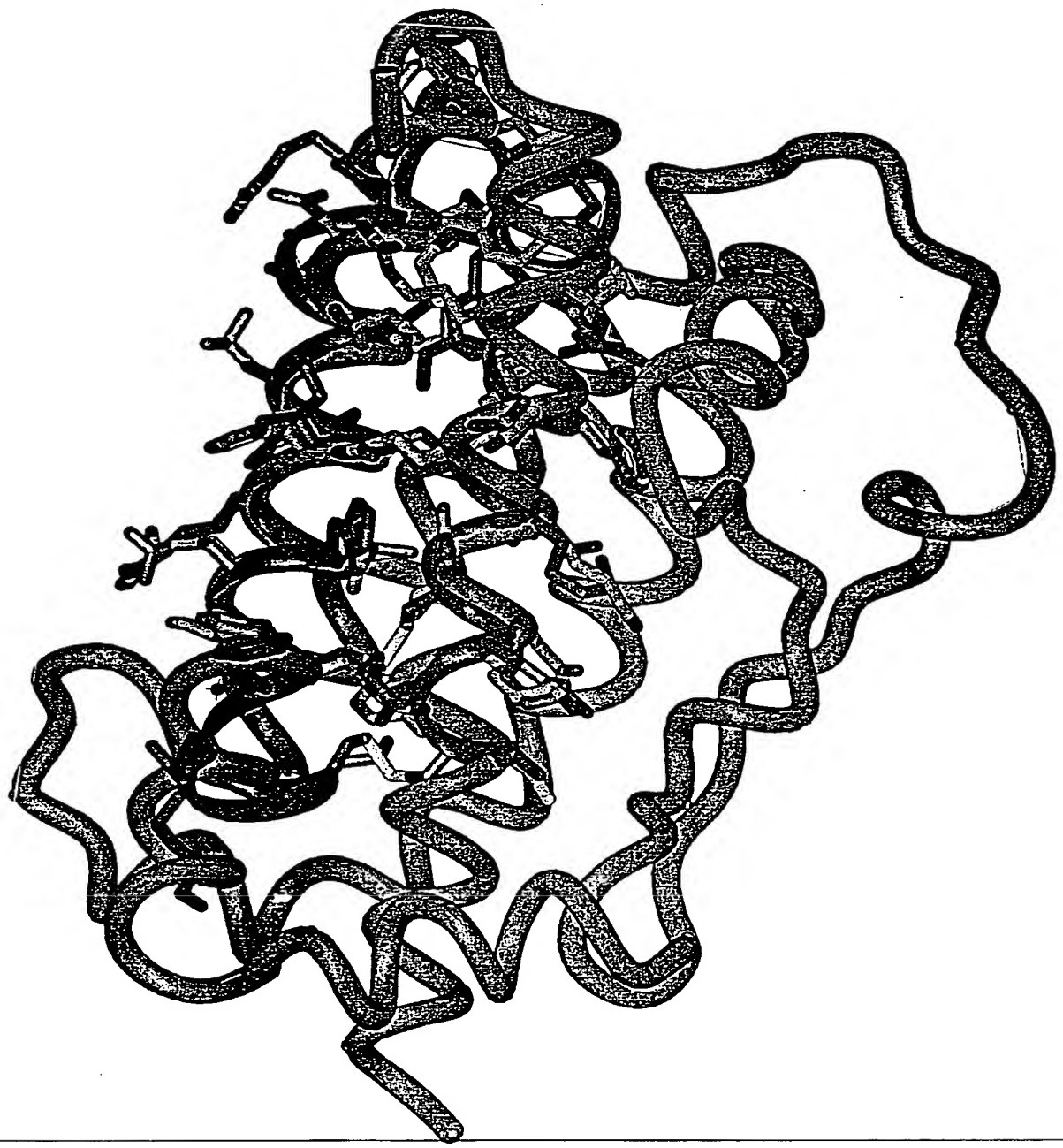


FIGURE 3

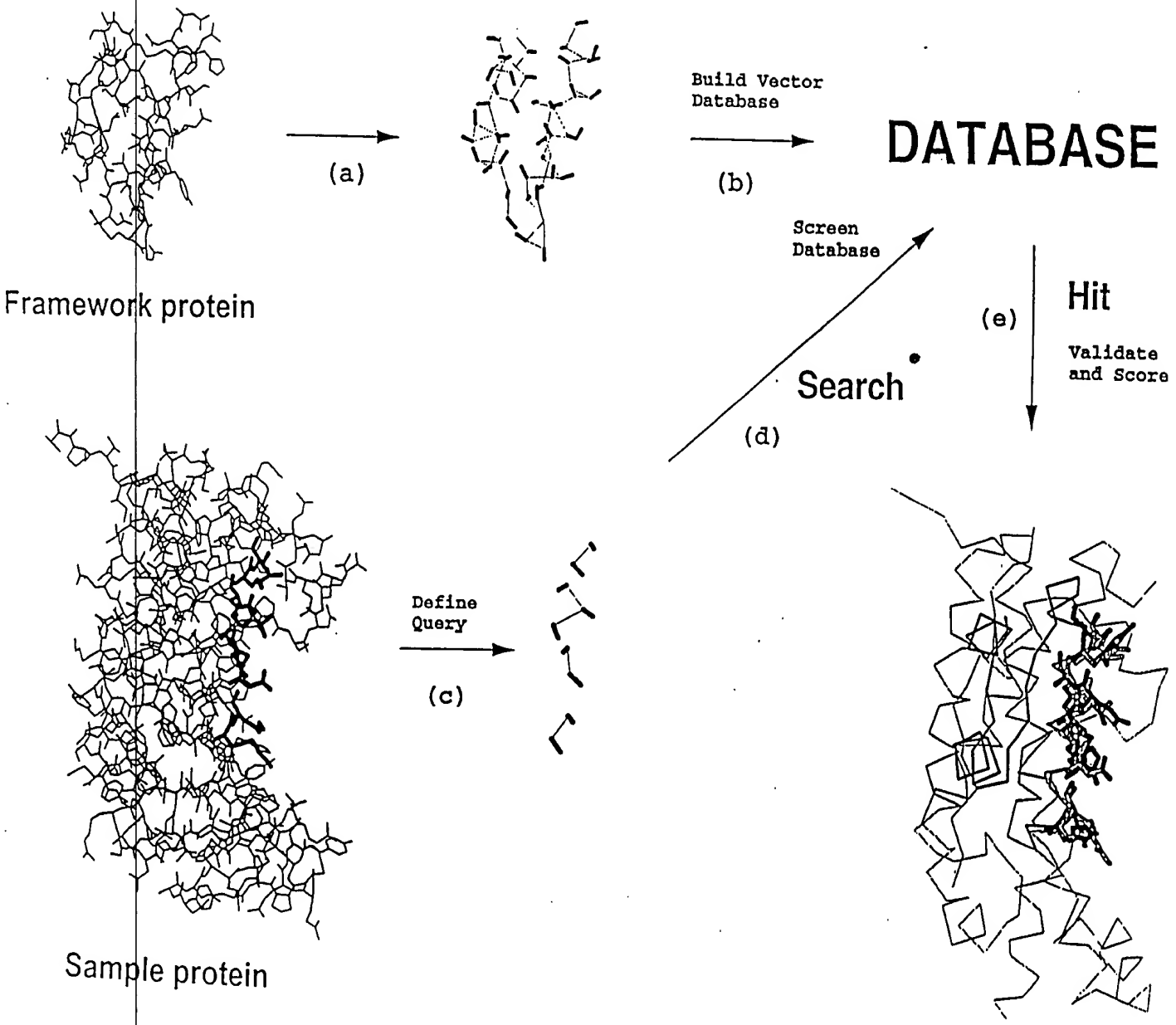

Framework protein
(a)
Build Vector Database
(b)
DATABASE
Screen Database
Search
(d)
(e)
Hit
Validate and Score
Define Query
(c)
Sample protein

FIGURE 4

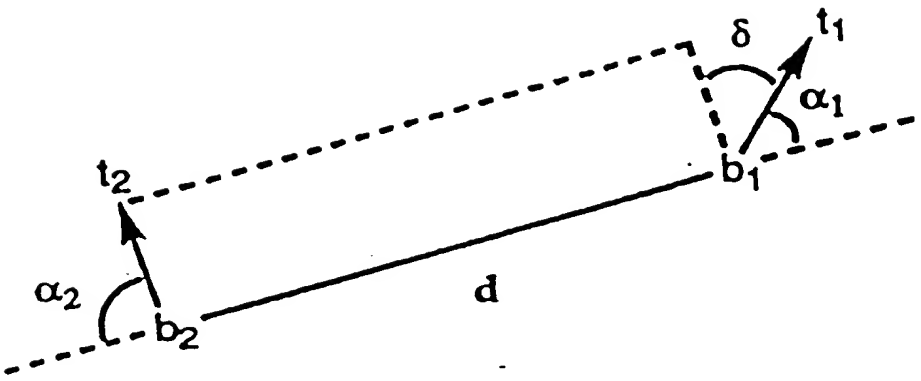


FIGURE 5

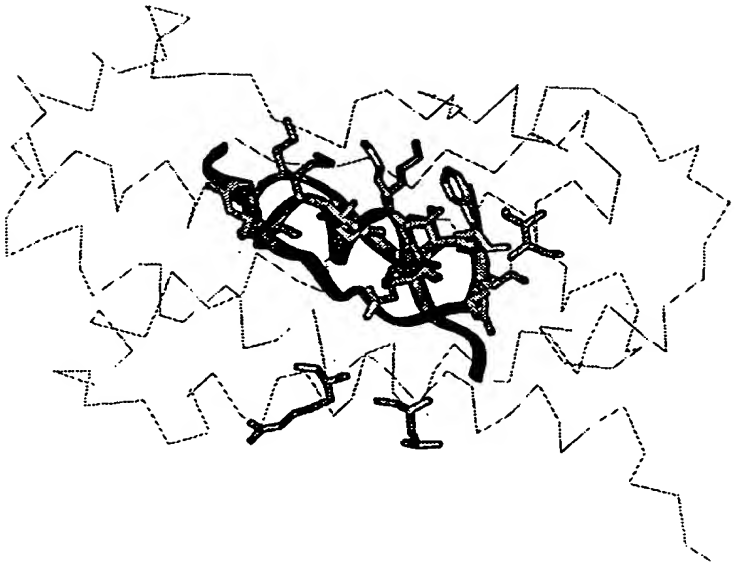


FIGURE 6

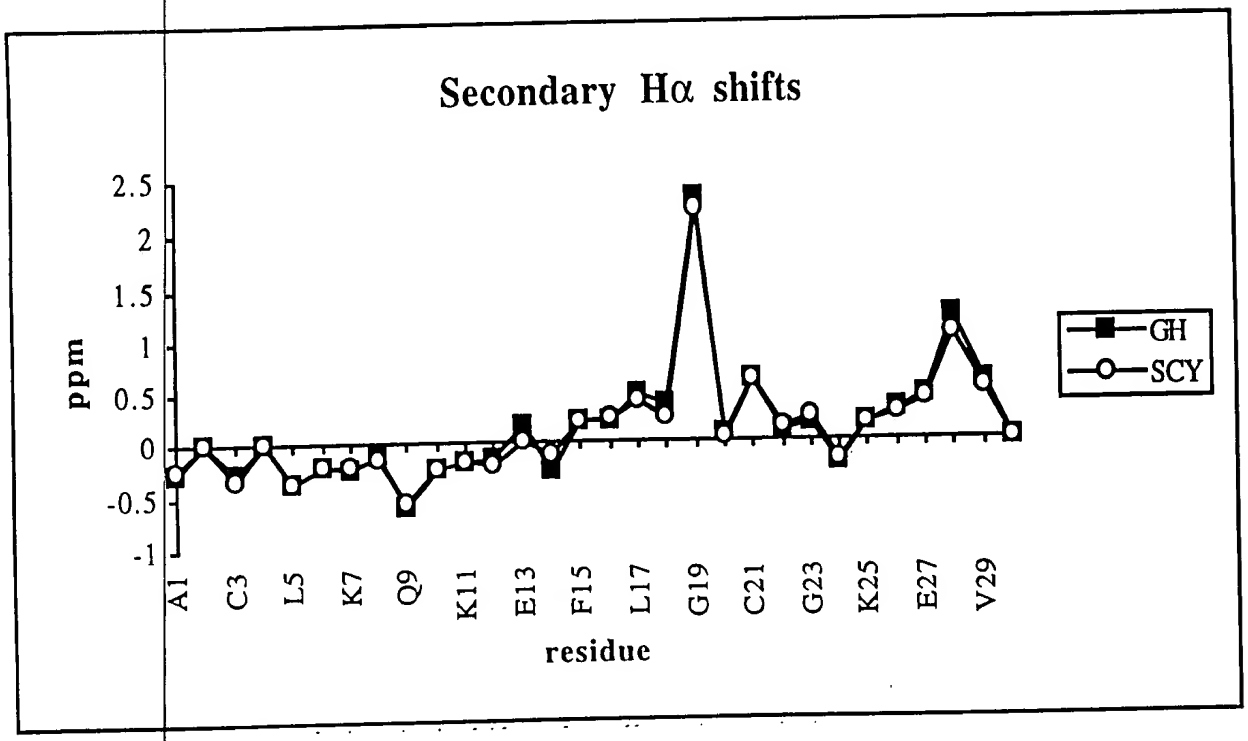

Secondary Hα shifts
2.5
2
1.5
1
0.5
0
-0.5
-1
ppm
A1
C3
L5
K7
Q9
K11
E13
F15
L17
G19
C21
G23
K25
E27
V29
residue
GH
SCY

FIGURE 7

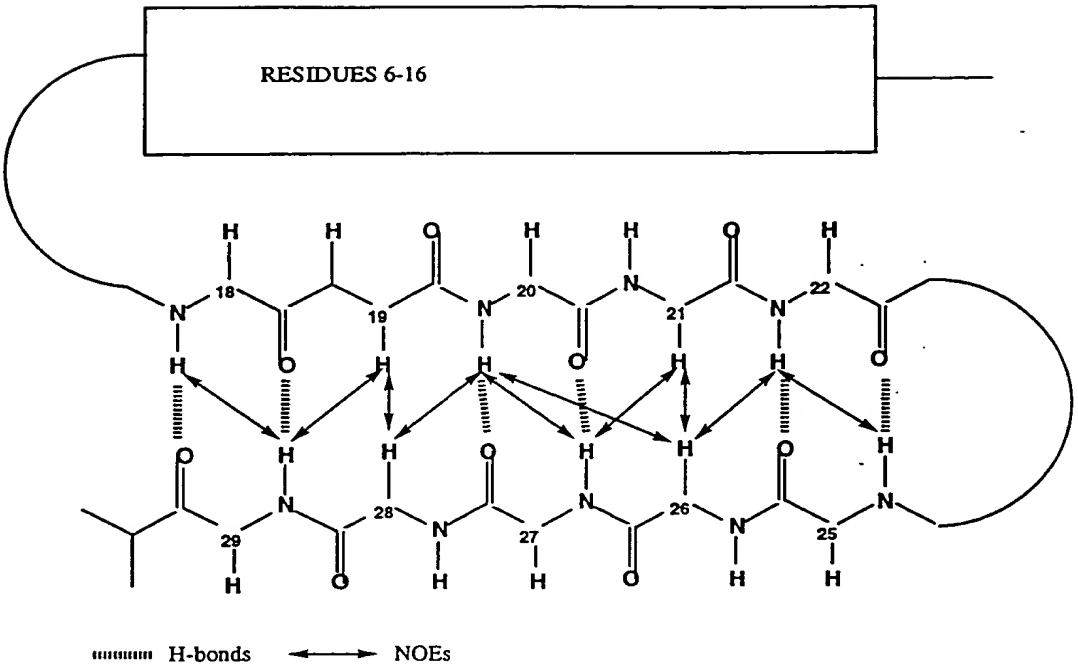


FIGURE 8

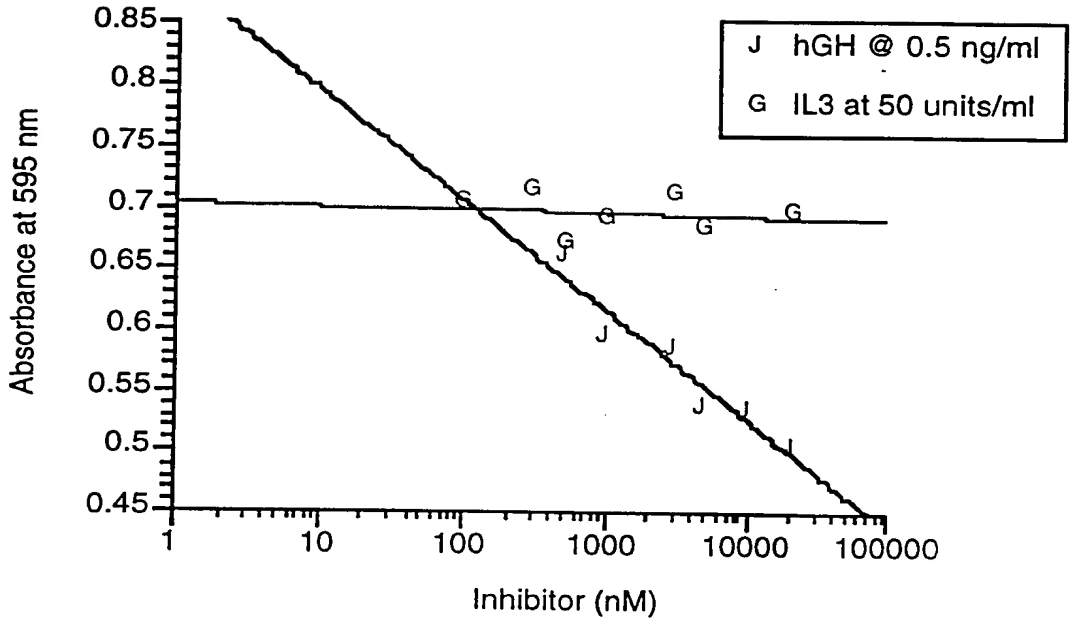
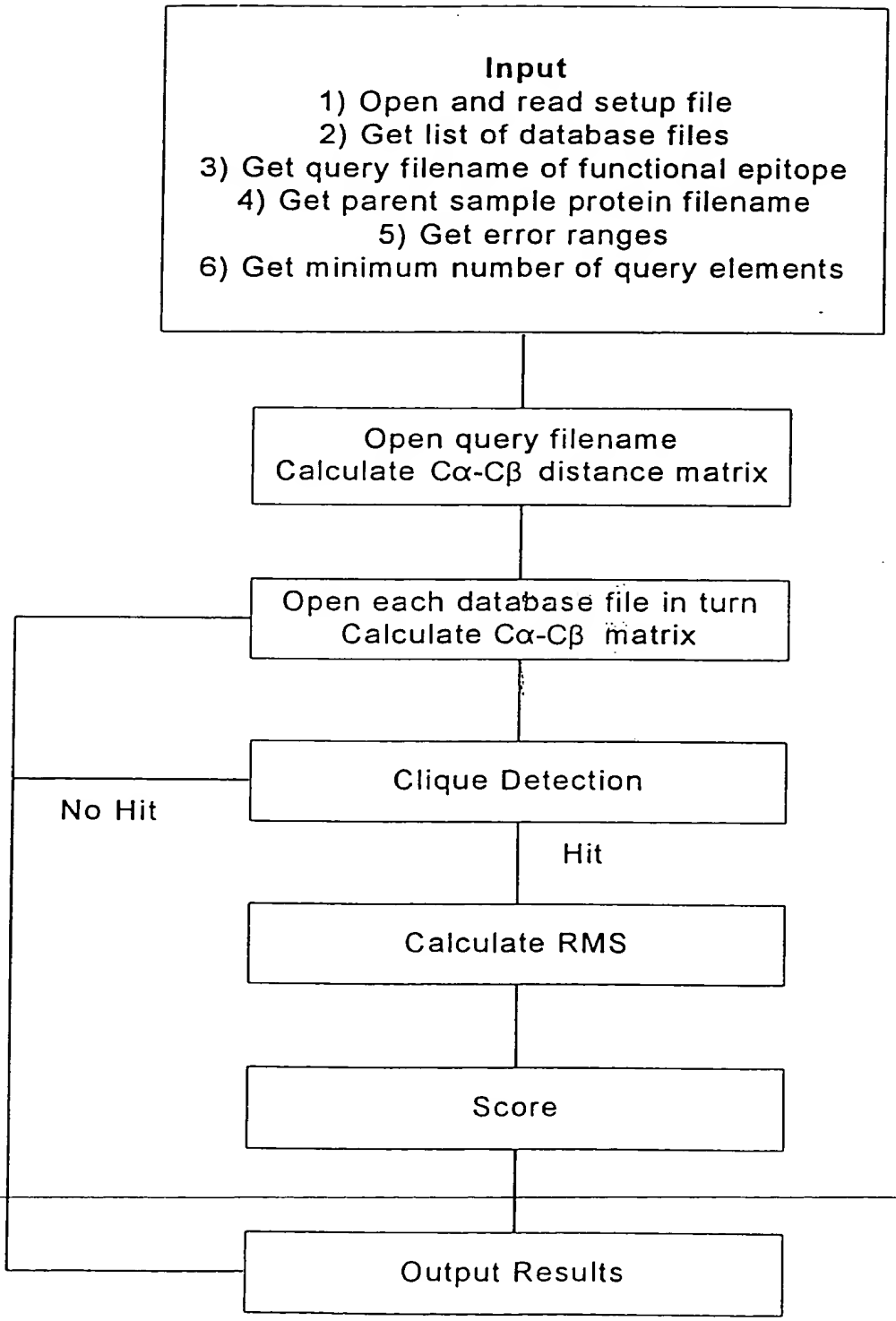
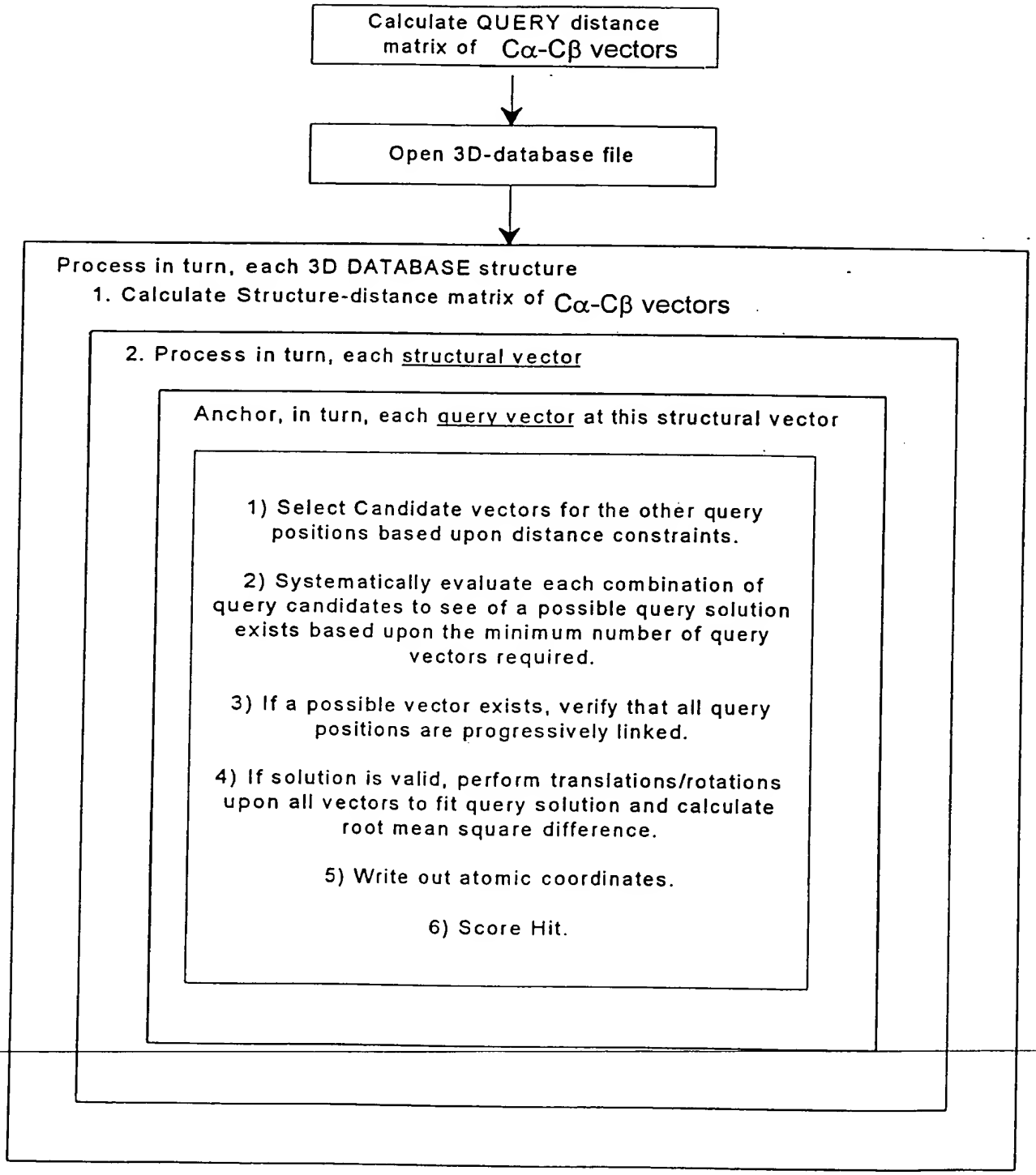

J hGH @ 0.5 ng/ml
G IL3 at 50 units/ml
Absorbance at 595 nm
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
1
10
100
1000
10000
100000
Inhibitor (nM)

FIGURE 9

Scheme A


Input
1) Open and read setup file
2) Get list of database files
3) Get query filename of functional epitope
4) Get parent sample protein filename
5) Get error ranges
6) Get minimum number of query elements
Open query filename
Calculate Cα-Cβ distance matrix
Open each database file in turn
Calculate Cα-Cβ matrix
Clique Detection
No Hit
Hit
Calculate RMS
Score
Output Results

Scheme B


Calculate QUERY distance matrix of Cα-Cβ vectors
Open 3D-database file
Process in turn, each 3D DATABASE structure
1. Calculate Structure-distance matrix of Cα-Cβ vectors
2. Process in turn, each structural vector
Anchor, in turn, each query vector at this structural vector
1) Select Candidate vectors for the other query positions based upon distance constraints.
2) Systematically evaluate each combination of query candidates to see of a possible query solution exists based upon the minimum number of query vectors required.
3) If a possible vector exists, verify that all query positions are progressively linked.
4) If solution is valid, perform translations/rotations upon all vectors to fit query solution and calculate root mean square difference.
5) Write out atomic coordinates.
6) Score Hit.

Scheme C

Automatic weekly job

Find likely new frameworks

1 Creates new sequentially named diectory for candidate structures (dirX)

2 Finds files created in last 7 days and filters the entries with small peptides (<70) and disulfides. Files are copied to directory dirX and clipped to one conformer. Source log file is created for InsightII

Manual visual check

1 Files are viewed in InsightII and reject files discarded

Database build

After manual visual check, run script./db_build in Search Database which performs the following functions:

1 Finds useful hits in directory dirX and copies these files from pdb database to Search Database.

2 Files are then split and renamed

3 The database is cleaned up

4 The original pdb files of the hits in pdb database are linked to Unique Database to produce an easily accessible record of all entries in the Search Database for characterisation etc.

5 The new entry list is appended to the MASTER_DB_LIST with the date.

Scheme D

(1) Gross steric filter: Manual visualisation of framework hit superimposed on the sample molecule in context of sample molecule's receptor if available. (discard hits with major steric clashes with receptor or hits that protrude beyond the surface of the ligand if receptor structure is not available

passed ↓ failed → DISCARD

(2) "Alanine scan": (Only if receptor structure is available.) Framework hit is reduced to polyalanine sequence, with identical backbone and Cb coordinates as original hit. Polyalanine molecule is "bump–checked" with the sample molecule's receptor to identify less obvious steric clashes with the receptor

passed ↓ failed → DISCARD or rectify(a)

(3) Framework hit mutated to sample molecule residues: Using the Biopolymer module of InsightII, the residues matching the sample molecule are mutated to the residue type of the sample molecule. Using the Search_compare module of InsightII, the sidechains of the mutated residues of the framework hit are flexibly fitted to the sidechains of their corresponding sample molecule residues to produce a theoretical 'Bioactive conformation'. Perform bump–check with receptor to identify unmutated sidechain steric clashes with receptor. Investigate mimicking unmatched functional residues of sample molecule with unnatural amino–acids

passed ↓ failed → DISCARD or rectify(b)

(4) Conformational stability of theoretical 'Bioactive conformation': The 'Bioactive conformation' from above is minimised in a forcefield. The rmsd of the backbone atoms of the minimised mutated framework hit from the minimised unmutated framework hit is calculated. If the rmsd<2.0 A, the conformation is considered accessible.

passed ↓ failed → DISCARD or rectify(c)

(5) Stability of fold: The mutated framework hit and the native framework hit are subject to molecular dynamics at 300K. The rmsd of minimised trajectory intermediates from the original conformer in each dynamics run are plotted. Unless there is a significantly greater drift in rmsd of the mutated relative to native framework hit, the fold is considered stable.

passed ↓ failed → DISCARD or rectify(c)

(6) Electrostatic similarity to target: Electostatic isocontour surfaces are generated for both mutated framework hit and sample molecule and compared for similarity. Electrostatic fields are mapped onto solvent accessible surface of mutated framework hit and sample molecule to compare electrostatic properties at the contact surface

passed ↓ failed → DISCARD or rectify(d)

Synthesise

Scheme D/2

Rectify(a): If the contact is with Cb of alanine, consider glycine replacement. If the contact is with backbone atoms and if the residues are in a terminal strand, consider truncation of sequence before the contact residue. Structural implications for truncations need to be considered. If the contact is to residues within a loop or imortant secondary structure, discard the hit.

Rectify(b): If there are steric clashes, consider the most conserved mutation that removes the bump.

Rectify(c): If the theoretical 'Bioactive conformation' is not stable, check the Ramachandran plot of the hit for residues in disallowed regions. Consider stabilizing structure with unnatural amino acids eg alpha–amino isobutyric acid in alpha helix motifs.

Rectify(d): Manipulate the electrostatic field of the framework hit via deletion of charge or introduction of charge using uncharged or charged residue mutations respectively. The distribution of charged residues in the sample molecule can be used as a guide for the placement of mutations in the framework hit. The structural implications of the mutation must also be considered.